Proceedings of the
7th Graz Brain-Computer Interface Conference 2017 DOI: 10.3217/978-3-85125-533-1-58

MULTI-TIMESCALE SPECTRA AS FEATURES FOR CONTINUOUS WORKLOAD ESTIMATION IN REALISTIC SETTINGS

D. Miklody[1], P. Moessmer[1], T. Dettmann[2], K. Klinkenberg[3], B. Blankertz[1]

[1]Department of Computer Science, Technische Universität Berlin, Berlin, Germany
[2]Federal Waterways Engineering and Research Institute, Karlsruhe, Germany
[3]K+S projects, Rangsdorf, Germany

E-mail: miklody@tu-berlin.de

ABSTRACT: In the real-time indication of cognitive workload in realistic settings, the main challenge in comparison to laboratory studies is the missing control of environmental variables of the participant. This introduces strong intrasubjectal/situational changes in EEG and makes most indicators built on one part of the data fail in another. We propose novel features reflecting the spatio-spectral brain state on multiple timescales as information for the workload indication. This allows the classifier to interpret the current brain state (estimated from the last 10 s) in reference to the slower changing background state (estimated from the last 10 mins). In a validation on an inland-waterway study, classification improves in a semi-realistic simulator setting from 19 % average error down to 13 %. In a corresponding real-world EEG measurement onboard a freight-ship, the improvement is less pronounced but workload estimation is possible.

INTRODUCTION

For the continuous indication of cognitive workload from electroencephalography (EEG), often averages of extracted features based on small time windows of seconds to minutes are used. The EEG features used for classification/indication are usually bandpass-filtered, spatially-filtered, spectral components, combinations of those or similar approaches[1,2,3].

While smaller windows bring the indication closer to real-time, bigger windows smooth the output and often reduce the influence of artefacts and thus improve classification accuracy[1].

While in laboratory settings cognitive workload indication can mostly be successfully performed, the transfer to realistic or semi-realistic environments remains a challenge. First of all, the participant is mostly more aware and concentrated on all of her/his actions as they have real-world effects. Also, the reduced setting of a laboratory or simulator might not elicit all the perceptual input of a real scenario and, thus, brain states might be more profound, distinct and situation-dependent in realistic situations. In addition, EEG experiments usually consist of many repetitions the data of which is then averaged to filter out unrelated signal components. This is mostly unfeasible in realistic scenarios.

With the aim to handle in particular the stronger changes in brain signals in realistic scenarios, we introduce a method that provides the classifier with the information about the recent past spectrum of the EEG. This leads to the possibility of filtering current changes out of ongoing activity.

MATERIALS AND METHODS

EEG recordings: The EEG was recorded with a BrainVision BrainAmp setup involving EasyCap ActiCaps System with 32 gel-based Ag/Ag-Cl electrodes and a sampling frequency of $f_s = 1kHz$. This amplifier setup guaranteed stable electrode impedances and high immunity to all kinds of noise and artefacts. All electrode impedances were ensured to be below 10 kOhms before recording the actual EEG.

Preprocessing & Artefact reduction: The raw EEG was low-pass filtered at 40 Hz with a Chebyshev-filter of order 10 and then downsampled to 100 Hz. After that, a PCA-based artefact reduction was performed [4]. To this end, all PCA components with a higher standard deviation than $70\,\mu V$ were removed from the original data. After this, spectrograms were built involving the Fast-Fourier-Transform (FFT) with a Hamming-window of length $500\,ms$.

Moving Average Spectrum: For moving average calculation, 2 different window-lengths were chosen: 1 min and 10 mins. These were calculated on the spectrograms with sampling frequency $f_s = 100\,Hz$.

Feature Selection: We used two different sets of feature vectors for comparison of our new approach to a common approach. From each, the spectrograms and the moving averages, 10s averages were built without temporal overlap. The first investigations

 DOI: 10.3217/978-3-85125-533-1-58

were done on a feature vector x_a only containing the averages over 10s-windows of the spectrum from 3 to 20 Hz in 1 Hz bins. For the moving average spectrum approach, the 3 feature vectors of 10 s, 1 min and 10 mins - each containing Frequency x Channels - were then combined to a joint feature vector x_b.

Classification: The feature vectors were then classified in a 2-class approach involving linear discriminant analysis (LDA) with regularized shrinkage [5]. The classes were 'high' and 'low' workload according to expert ratings about the difficulty of the sailing in an inland waterway cargo ship task as described below in the section *Class labels*.

Validation: To estimate the performance of the classifier/indicator on unseen data, a cross-validation was performed with a chronological block-wise sampling. As error measures we decided for the class-wise normalized loss and the Area under Curve (AUC) of the receiver operator characteristics of the classifier output [6]. The AUC was inspected to see how well the classifier separates classes and could perform under a decision boundary shift.

Scalp Topographies: To investigate what features the classifier had actually picked, the resulting weight vectors were transformed to scalp topographies A by multiplication with the covariances of the features [7]:

$$A \propto \Sigma_X w$$

where Σ_X is the covariance matrix of the features X and w the LDA weight vector. For the classifier weights, the classifiers and the covariance were calculated on the whole dataset.

Experiments: Professional captains performed the passage of different bridges and one lock in a reference section between two locks on the river Main near Würzburg in Germany. Three captains performed several repetitions of the two bridge passages in the simulator of the Federal Waterways Engineering and Research Institute (*German* Bundesamt für Wasserbau BAW) in Karlsruhe with different water levels and travel directions. Additionally, another captain's EEG was measured while passing both locations and the part connecting them on a routine travel with the same cargo ship as in the simulator. The ship was a 186 m long push tow of 2 units (GMS 'Odeon' and GLS 'Elly').

Class labels: The labels for the workload classes "high" and "low" were chosen under consideration of expert ratings. The bridge passages under investigation were described by experts to be difficult. We chose the last 3 mins before the bridge passages in the *simulator* as "high" workload. In order to have a low workload phase to contrast this against, we chose the time window 1 min after the bridge passage as the ship was very long and took time to pass it. Also, a period of 15s after the start of each run until 4 mins after the start was used as a "low" workload time window. In the *on-board* scenario we chose longer periods of 7 mins in a similar scheme but added intermediate low workload phases to increase the available amount of data to the classifier and improve sampling over time.

RESULTS

Simulator: The classification in the simulator was in general successful with a class-wise average classification loss of 19 % for our common 10 s-window approach X_a and 13 % for the moving average based feature vector x_b, which can be seen in Table 1.

Table 1: Classification results

Subject	10s average	10s+1min+10min
SF1	26% (0.16)	19% (0.09)
SF2	13% (0.06)	11% (0.03)
SF3	18% (0.09)	10% (0.03)
Average	19% (0.10)	13% (0.05)
SF4 on-board	43% (0.41)	41% (0.34)

Class-wise normalized loss with AUC in parentheses.

We found spatial topographies that suggest neural origin as main features for classification in all participants. As we were operating in a semi-realistic environment with multiple angles of vision and the captains moving relatively freely in their chairs, eye movement, blinking, muscle and general movement artefacts were inevitable. PCA-based artefact reduction removed the strongest components which were mainly contributed to eye movements. But, as can be seen in Figure 1-4, some artefact components were most probably left in the data, discriminative of the workload and thus were used by the classifier. Strong local patterns from peripheral channels like frontal or occipital electrodes are likely to be artefacts in form of e.g. eye artefacts of SF1 or neck muscle artefacts of SF4. Still, the topographies for both x_a and x_b consisted mostly of central patterns, especially above 7 Hz. Interestingly, the classifier knowing the recent past had mostly weighted the 10 s average most strongly and often contrasted it against the recent past in form of 1 min and 10 min averages.

 DOI: 10.3217/978-3-85125-533-1-58

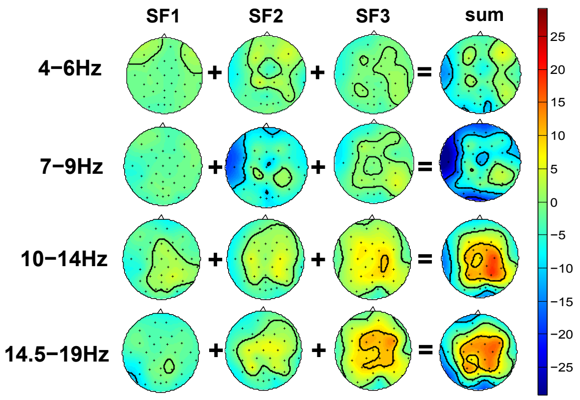

Figure 1: The scalp topographies of the classifier for feature vector x_a only containing 10 s average spectra. A positive sign represents positive weight for the class 'high' (the sign of the classifier is 'high'-'low').

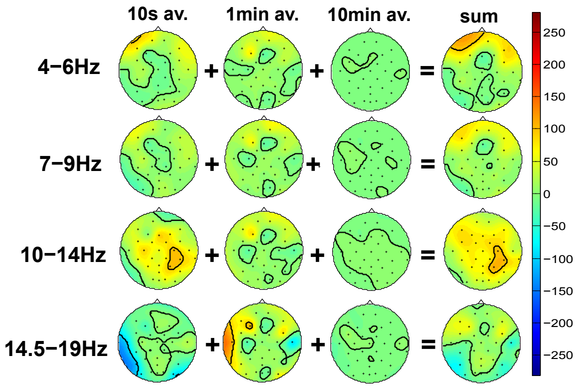

Figure 2: The scalp topographies for subject SF1 of the classifier for feature vector x_b containing 10 s, 1 min and 10 min average spectra.

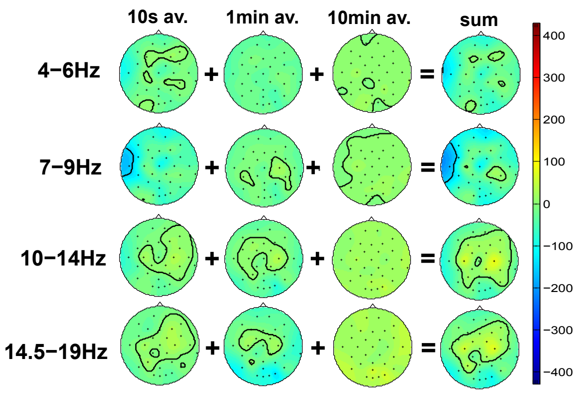

Figure 3: The scalp topographies for subject SF2 of the classifier for feature vector x_b containing 10 s, 1 min and 10 min average spectra.

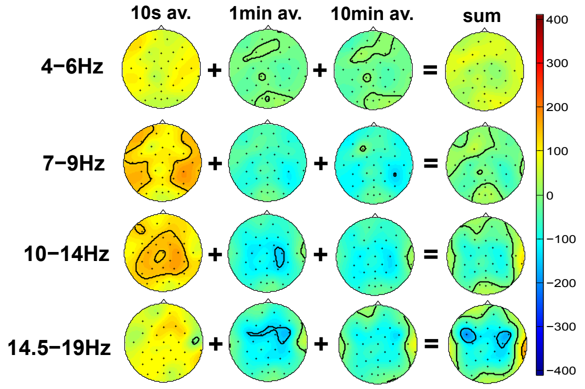

Figure 4: The scalp topographies for subject SF3 of the classifier for feature vector x_b containing 10 s, 1 min and 10 min average spectra.

In Figure 5 the Grand Average of the classifier output from a within subject block-wise cross-validation over the three subjects is shown on a map of the river Main. The classifier output interpreted as a workload indicator was in particular 'high' just in front of the bridges with an overall smooth increase and decrease.

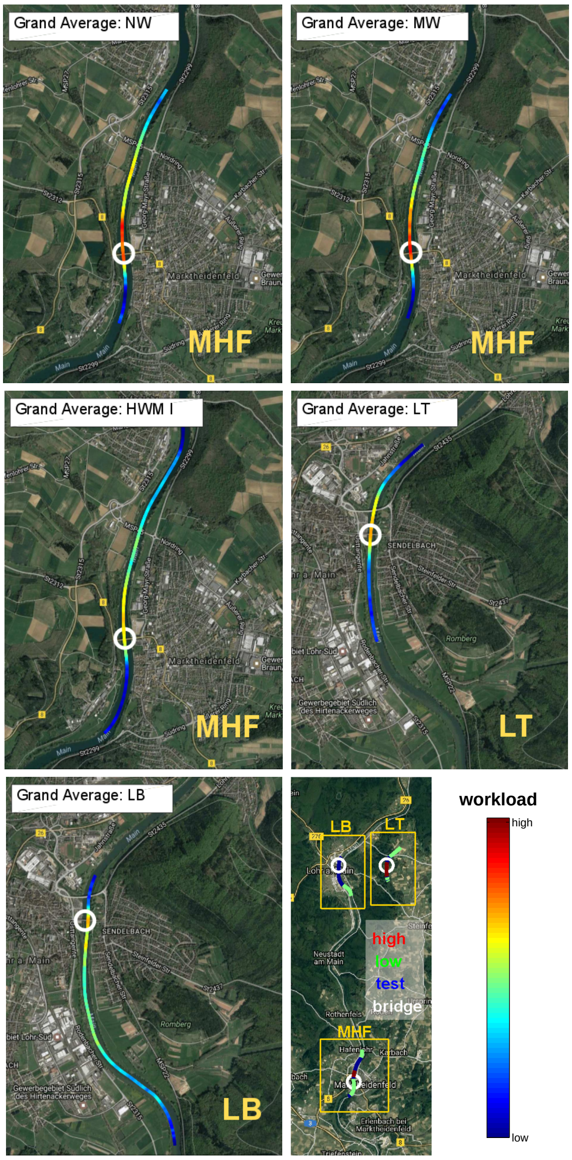

Figure 5: The workload indicator for the different scenarios in the simulator from multi-timescale feature vector x_b. The circles depict the position of the bridges. On the bottom right is an overview of the area with the two bridges and the labeling scheme. The abbreviations are as follows: *NW* low water bridge 1 - *MW* normal water bridge 1 - *HWM I* high water mark I bridge 1- *MHF* location bridge 1 - *LT* downstream bridge 2 - *LB* upstream bridge 2. *LT* is shifted horizontally on the overview map to fit into one picture.

On-Board: In the on-board data, classification worked less well in general. Classification could be performed with a class-wise averaged loss of 41% (AUC 0.34) with feature vector x_b, while the 10 s average based x_a performed with 43% (AUC 0.41), which can be examined in Table 1. The output of the multi-timescale feature vector x_b was much smoother, than that of x_a. The classifier was trained on much less samples and, also, more artefacts were visible in the spectra.

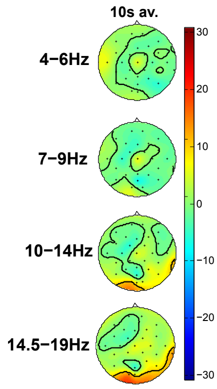

Figure 7: The scalp topography of the classifier for for subject SF4 with the 10 s only feature vector x_a. A positive sign represents positive weight for the class 'high' (the sign of the classifier is 'high'-'low').

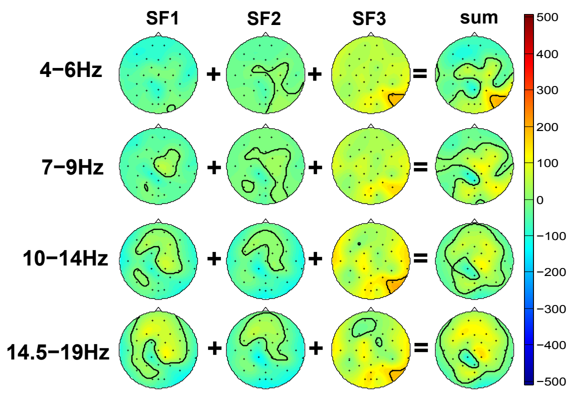

Figure 8: The scalp topography of the classifier for subject SF4 with the 10 s-1min-10 min feature vector x_b.

DOI: 10.3217/978-3-85125-533-1-58

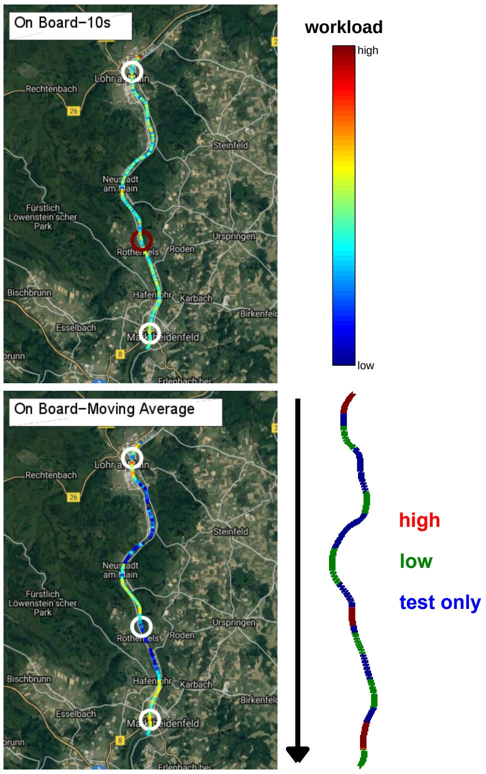

Figure 9: The workload indicator of the on-board experiment for the 10 s feature vector x_a and the multi-timescale feature vector x_b on the river Main. The upper most circle depicts bridge 2, the one in the middle the lock and the lower circle bridge 1.

DISCUSSION

Simulator data: The simulator data can be classified with a high average accuracy of 80-90 % for all of the three subjects. Scalp topographies might still include systematic movement artefacts but the connected motor activity in the brain also appears discriminative. Activity over primary sensory and motor areas of the hands could have lead to the slightly lateralized central activity in the alpha/mu-band. The workload indication within the labels was successful with high accuracies and also the overall picture of workload indication in the unmarked time spans was reasonable. Workload was "high" only in turns or close to bridges while staying "low" in between.

On-Board: Classification of the on-board data seems more challenging in general but also the much smaller amount of data has to be taken into account. Additionally, the investigation of behavior lead to very different impressions for the different time points of the trip, that were not solely connected to the bridge and lock passages. Additional communication took place, unplanned vessels crossed the way and more movement was necessary for the captain as the bridge was bigger. One big point was that the main display was mounted to the left of the captain, at which he seemed to look more frequently during sailing-induced higher workload. Accordingly, the main challenge with the on-board data of this subject seems to be the systematic motor activity connected to the observation of the display to the left of the captain. This could also be found in the scalp topographies connected to the classifier weights, as the neck muscles could have lead to the strong signals from the right occipital electrodes used for classification.

Moving average classifier: The moving average classifier improves the classification accuracy on the simulator data by 6 % on average. The scalp topographies suggest that mainly the actual 10 s-data x_a was effectively used and the 1 min and 10 mins moving averages were only slightly used, but improved the results. This supports our hypothesis that supplying the classifier with the recent past makes it more robust to experimentally unconnected temporal changes commonly exhibited in the EEG within single subjects. Compared to the simulations, more of the 10 min average was used in the on-board classification in general. All kinds of artefacts seem more pronounced for this realistic scenario. Less data is available and the class labels are probably more noisy because many more latent variables not related to the bridge and lock passage have an effect on the cognitive workload. The classifier performed slightly better than the 10 s only feature vector x_a-based.

Classifier Scalp Topographies: The topographies show in general a main effect of brain activity while more predominant artefacts are also discriminative for the classification compared to most laboratory studies. Different movement patterns and viewing directions are most probably significant of the class label, as in the high-workload phase most captains seemed to steer more and also watch the displays more frequently. This is inevitable but, additionally, the results are not fully homogeneous over subjects.

CONCLUSION

In a series of linked simulator and on-board experiments on the cognitive workload of professional inland waterway captains, we successfully performed classifications involving LDA on multi-timescale spectral features. The classifier output generalizes between situations and can be used as a linear workload indicator within our data. The workload indication in semi-realistic and realistic scenarios seems possible in general. More challenges might be raised by the step from the simulator onto the board of a real ship, as performance drops drastically here. Still, our approach involving the data of the most recent past seems to perform slightly better than a simple 10 s window feature vector in the on-board scenario and improves the results from the simulator. Sadly, the simulator data of subject SF4 could not be recorded to compare and transfer

 DOI: 10.3217/978-3-85125-533-1-58

the simulator results from this subject to the on-board measurements. It was planned to also take the on-board captain to the experiments in the simulator but, due to timing issues, this was not possible. This is planned for the future and could help to measure the transferability of the indicator extracted from simulator data to more realistic on-board scenarios. Also, the amount of training data would be increased by this. As this was a simple pilot study only involving three subjects in the simulator and one on-board, results have to be carefully interpreted with this limitation in mind. Still, the differences in the results for different subjects, in particular for the topographies, could be related to different individual strategies. For example, one captain may correct his course more in the "high"-workload phase, while another might be more precise and concentrated and thus correcting his course less. Also, one captain may solve the tasks more visually while another is more focused on his sensory input and motor actions which leads to different brain activity patterns. The multi-timescale spectral feature based workload indicator has to be further investigated but first results show relevance to realistic EEG experiments.

REFERENCES

[1] Miklody D, Uitterhoeve WM, v Heel D, Klinkenberg K, Blankertz B. Maritime Cognitive Workload Assessment, in Symbiotic Interaction, Springer International Publishing, 9961, 2016.

[2] Kohlmorgen J, Dornhege G, Braun M, Blankertz B, Müller KR, Curio G, Hagemann K, Bruns A, Schrauf M, Kincses W. Improving human performance in a real operating environment through real-time mental workload detection, in Toward Brain-Computer Interfacing, MIT press, 2007.

[3] Brouwer AM, Hogervorst MA, Van Erp JB, Heffelaar T, Zimmerman PH, Oostenveld R. Estimating workload using EEG spectral power and ERPs in the n-back task, in Journal of neural engineering 2012, 9(4), 045008.

[4] Casarotto S, Bianchi AM, Cerutti S, Chiarenza GA.. Principal component analysis for reduction of ocular artefacts in event-related potentials of normal and dyslexic children, in Clinical Nneurophysiology 2004, 115(3), 609-619.

[5] Vidaurre C, Krämer N, Blankertz B,Schlögl A. Time domain parameters as a feature for EEG-based brain–computer interfaces, in Neural Networks 2009, 22(9), 1313-1319.

[6] Fawcett T. An introduction to ROC analysis, in Pattern Recognit. Lett. 2006, 27, 861–74.

[7] Haufe S, Meinecke F, Görgen K, Dähne S, Haynes JD, Blankertz B, Bießmann F. On the interpretation of weight vectors of linear models in multivariate neuroimaging, in NeuroImage 2014; 87: 96-110.